REPORT DOCUMENTATION PAGE

1

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
19-09-2014	Final Report	19-Jul-2010 - 18-Jul-2013

4. TITLE AND SUBTITLE

Final Report: Developing Unconstrained Methods for Enzyme Evolution

5a. CONTRACT NUMBER

W911NF-10-1-0298

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

611102

6. AUTHORS

Korch S.B., Stomel J.M., Leon M.A., Hamada M.A., Stevenson C.R., Simpson B.W., Gujulla S.K., Chaput J.C.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

Arizona State University
PO Box 876011

Tempe, AZ 85287 -6011

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

57913-LS.2

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

A grand challenge in synthetic biology is to create artificial enzymes with catalytic activities similar to natural enzymes. Although several protein enzymes have been developed by computational design and protein evolution methods, the generation of efficient enzymes remains a difficult problem. In this sponsored project we are examining the question of why modern protein engineering methods fail to produce catalytically efficient enzymes. This study has broad application in many technologies from chemical synthesis to human health and the environment. Our work centers around the notion that de

15. SUBJECT TERMS

de novo protein evolution

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			John Chaput
UU	UU	UU	UU		19b. TELEPHONE NUMBER 480-727-0392

Report Title

Final Report: Developing Unconstrained Methods for Enzyme Evolution

ABSTRACT

A grand challenge in synthetic biology is to create artificial enzymes with catalytic activities similar to natural enzymes. Although several protein enzymes have been developed by computational design and protein evolution methods, the generation of efficient enzymes remains a difficult problem. In this sponsored project we are examining the question of why modern protein engineering methods fail to produce catalytically efficient enzymes. This study has broad application in many technologies from chemical synthesis to human health and the environment. Our work centers around the notion that de
novo evolved proteins represent better starting points for catalyst development than natural proteins,because unlike natural proteins, synthetic proteins are not biased by a complex, largely unknown evolutionary history. To test this hypothesis, we are attempting to evolve a de novo generated ATP
binding protein into a larger protein structure with improved ATPase activity. This project examines two important questions: (1) to what extent can protein evolution methods be used to transform small protein folds into larger globular structures; and (2) what physical constraints limit the evolution of synthetic protein enzymes? In year one, we constructed several mRNA display libraries that contained random regions of 20-, 40-, and 80- amino acids added to the C-terminus of our starting ATP binding
protein scaffold. During year one, we also developed a functional reporter assay for protein folding. In year two, we used mRNA display to evolve a series of enlarged ATP-dependent protein folds. Using our protein-folding assay, we have identified several clones that remain stably folded and soluble in E. coli. We are now in the process of characterizing the expression and solubility properties of the most promising candidate proteins.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received Book

TOTAL:

Received	Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Ms. Stevenson was a recipient of the ASU Merit Scholarship and US Army Undergraduate Summer Research Fellowship

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Andrew Larsen	0.50	
Bing Jiang	1.00	
FTE Equivalent:	**1.50**	
Total Number:	**2**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Sunil Kumar	1.00
FTE Equivalent:	**1.00**
Total Number:	**1**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
John C. Chaput	0.10	No
FTE Equivalent:	**0.10**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED	Discipline
Christine Stevenson	0.10	Biochemistry
Brent Simpson	0.10	Cellular Biology
Ayush Gupta	0.10	Biochemistry
FTE Equivalent:	**0.30**	
Total Number:	**3**	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	3.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:	3.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:	3.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):	3.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:	1.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	3.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME	
Bing Jiang	
Total Number:	**1**

Names of other research staff

NAME	PERCENT_SUPPORTED
William Selleck	0.50
FTE Equivalent:	**0.50**
Total Number:	**1**

Sub Contractors (DD882)

Inventions (DD882)

This grant enabled us to make considerable progress toward our long-term goal of understanding the physical constraints that limit protein enzyme development.

[1] mRNA Display selection for large globular proteins capable of binding ATP.We have completed seven rounds of in vitro selection and amplification using mRNA display to isolate synthetic proteins that can fold into structures that recognize ATP with high affinity and specificity. Our
starting library (designed and constructed in year 1) contained a common N-terminal de novo evolved ATP binding protein followed by an unbiased random region of 80-amino acids, which was designed to expand the original protein fold by one or more new protein domains. To favor the isolation of stably
folded proteins, we included guanidine hydrochloride (GuHCl) in the selection buffer to remove weak or poorly folded proteins from the pool. The concentration of GuHCl was gradually increased over the course of the selection from 0 to 2.0 M. We monitored the selection progress using a cell-based fluorescence assay (see below), which indicated that the relative abundance of stably folded proteins increased significantly from round 0 to round 6 and plateaued in round 7. To determine the diversity of sequences that remained in the pool after seven rounds of in vitro selection and amplification, we cloned 100 representative sequences. As expected, all of the sequences share a common N-terminal sequence that defines the boundary of the parent protein. The C-terminal region of each protein was unique, which may indicate that there are many distinct solutions to the problem of how larger protein structures can emerge from smaller protein domains.

[2] Cell-based fluorescence screen to monitor selection progress and identify soluble protein variants. We have developed a fluorescence-activated cell-sorting (FACS) assay to monitor the selection
progress and identify well-folded and soluble protein variants present in the output of our mRNA display library. The screen is based on a green fluorescent protein (GFP) reporter assay developed by Terwilliger and Waldo (Nature Biotech. 1999, 17, 691) that relies on the formation of a productive
folding conformation of the upstream analyte when expressed as a fusion protein in Escherichia coli. In year 1, we designed and constructed the GFP report vector. This was necessary as we were unable to procure this vector from the Waldo lab, despite several attempts to do so. Using the GFP reporter vector, we screened populations of sequences from each round of selection for protein folding in Escherichia coli. FACS analysis revealed that the starting library exhibited very low fluorescence, similar to the negative (empty vector) control. This is expected as randomly chosen
sequences should not fold into stable structures. Analysis of sequences taken from each round of selection showed a gradual increase in florescence, indicating that the selection successfully enriched for proteins with improved protein folding stability.To evaluate the selection output, we cloned five representative sequences from round 7 into the GFP
reporter and assayed each sequence for protein folding in Escherichia coli. Two of the five clones, gave a single population of highly fluorescent cells equivalent to our positive control, which contained only parent protein. The other three sequences gave mixed populations of low and high fluorescent
cells, indicating that these sequences were adopting partially folded structures.Encouraged by our FACS analysis, we cloned 64 sequences into the GFP vector and performed a 96-well fluorescence-based assay to identify a subset of sequences from round 7 that were properly folded and soluble. This assay revealed that 18 out of 64 sequences have high relative fluorescence that is reproducible (triplicate). We sequenced the 18 clones and identified 12 sequences with intact open reading frames. These 12 sequences were examined in a standard expression vector (below).

[3] E. coli expression analysis of randomly selected protein clones from round 7.In addition to the 12 sequences identified in our 96-well protein-folding assay, we chose an additional 23 sequences from round 7 for protein expression analysis. All 35 sequences were cloned into a protein expression vector containing an N-terminal maltose binding protein (MBP) and expressed as C-terminal protein fusions of MBP. The 35 ATP binding protein fusions were expressed in E. coli and protein expression was monitored after purification on amylose affinity columns. This analysis led us
to identify 4 promising clones that remained soluble as MBP-protein fusions at room temperature for several days. Three of the four proteins yield visible amounts of soluble free protein after cleavage of the fusion protein with thrombin.

[4] Expression optimization by mutagenesis.
To improve protein expression and recovery of the free ATP binding proteins, we made a series of Cterminal truncations to remove amino acid residues that were deleterious. This is a common approach to improving protein solubility and expression. In all three cases, we observed improved expression and solubility when 12 amino acid residues were deleted from the C-terminus.

[5] E. coli expression and purification of three candidate proteins with high solubility. We are currently optimizing the expression and purification of the three protein candidates so that we can obtain sufficient amounts of protein for biophysical characterization. We are particularly interested in performing an HSQC NMR analysis on each protein to determine the extent to which each protein adopts a discrete protein fold. We are hopeful that at least one of our three proteins is sufficiently well folded to move forward with structure determination studies by NMR and X-ray crystallography.

[6] Protein Characterization by NMR and X-ray crystallography. The top performing candidate protein identified in our protein solubility screen was selected for structure determination studies by NMR and X-ray crystallography. After considerable effort, conditions were identified that allowed us to obtain milligram quantities of highly pure protein in

sufficient concentrations (20 mg/mL) for protein crystallization. However, despite extensive efforts, we were unable to identify suitable crystallization conditions. Because of this result, we shifted our efforts
to structure determination by solution NMR. Despite high expression in standard LB media, the protein was extremely difficult to express in minimal media with N-15 labeled ammonia. After several months of screening, we finally identified conditions that allowed us to obtain labeled protein in
purified form. HSQC NMR experiments revealed that the evolutionary optimized protein likely adopts a molten globular structure. This unfortunate outcome led us to redesign our library and selection strategy and new efforts are underway to obtain a novel de novo evolved protein with an expanded
domain structure.

Technology Transfer

a. Abstract

A grand challenge in synthetic biology is to create artificial enzymes with catalytic activities similar to natural enzymes. Although several protein enzymes have been developed by computational design and protein evolution methods, the generation of efficient enzymes remains a difficult problem. In this sponsored project we are examining the question of why modern protein engineering methods fail to produce catalytically efficient enzymes. This study has broad application in many technologies from chemical synthesis to human health and the environment. Our work centers around the notion that de novo evolved proteins represent better starting points for catalyst development than natural proteins, because unlike natural proteins, synthetic proteins are not biased by a complex, largely unknown evolutionary history. To test this hypothesis, we are attempting to evolve a de novo generated ATP binding protein into a larger protein structure with improved ATPase activity. This project examines two important questions: (1) to what extent can protein evolution methods be used to transform small protein folds into larger globular structures; and (2) what physical constraints limit the evolution of synthetic protein enzymes? In year one, we constructed several mRNA display libraries that contained random regions of 20-, 40-, and 80- amino acids added to the C-terminus of our starting ATP binding protein scaffold. During year one, we also developed a functional reporter assay for protein folding. In year two, we used mRNA display to evolve a series of enlarged ATP-dependent protein folds. Using our protein-folding assay, we have identified several clones that remain stably folded and soluble in E. coli. We are now in the process of characterizing the expression and solubility properties of the most promising candidate proteins.

b. Publications

1. Korch, S.B., Stomel, J.M., Leon, M.A., Hamada, M.A., Stevenson, C.R., Simpson, B.W., Gujulla, S.K., and **Chaput, J.C.*** 2013. ATP sequestration by a synthetic ATP-binding protein leads to novel phenotypic changes in *Escherichia coli*. ACS Chemical Biology 8, 451-456.

c. Student Support

Dr. Sunil Gujulla: Dr. Kumar was a postdoctoral student in the Chaput lab. On July 1, 2012 he accepted a research faculty position in the Republic of Singapore. While in the lab, he worked to develop a GFP-based fluorescent screen for evaluating the solubility and expression properties of protein libraries containing 20-, 40-, and 80-random amino acid positions at the C-terminus of the native protein fold. He also worked closely with Ms. Jiang to screen and characterize individual clones identified after 7 rounds of in vitro selection and amplification.

Mr. Andrew Larsen: Mr. Larsen is a Ph.D. student in the Biological Design Program in the Biodesign Institute. As part of his Ph.D. thesis, he designed and constructed unbiased DNA libraries encoding 20-, 40-, and 80-contiguous random amino acid positions. These libraries were designed so that they could be ligated onto the C-terminus of our synthetic ATP binding protein. Mr. Larsen is currently assisting with sequence analysis and protein expression.

Ms. Bing Jiang: Ms. Jiang is a Ph.D. student in the Department of Chemistry and Biochemistry. As part of her Ph.D. thesis, she developed an mRNA display selection strategy to evolve novel ATPase and kinase enzymes from pools of random sequences. She performed 7 rounds of in vitro selection and amplification. Her characterization revealed that most of the proteins isolated from her selection adopted molten globular structures.

Ms. Christine Stevenson: Ms. Stevenson is an undergraduate student in the Barrett Honor's College at ASU. In 2011, she received a U.S. Army Undergraduate Research Fellowship. For her fellowship, she worked with Dr. Kumar and Ms. Jiang to develop a GFP reporter system for protein folding. She assisted throughout the year with protein expression and purification.

Ms. Shivani Kothari: Ms. Kothari is a high school student at BASSIS Scottsdale. In 2011, she received a U.S. Army High School Summer Research Fellowship. Ms. Kothari spent the summer learning basic techniques in DNA cloning and microscopy.

Mr. Brent Simpson: Mr. Simpson was an undergraduate student in the Department of Chemistry and Biochemistry. Mr. Simpson worked closely with Mr. Larsen to assist in vector design and protein expression. He also worked closely with Shaleen Korsh to study the effects of our de novo evolved proteins in living bacteria cells.

Mr. Ayush Gupta: Mr. Gupta is an undergraduate student at the University of California Berkeley. He spent the 2012 summer working with Dr. Kumar and Mr. Larsen on protein expression protocols.

Mr. Will Selleck: Mr. Selleck joined the Chaput lab in August 2012. Mr. Selleck is an expert in protein expression, purification, and automation. He will replace Dr. Kumar and work closely with Ms. Jiang on protein expression, purification, and structure determination studies.

d. Student Metrics

Christine Stevenson, Brent Simpson, and Ayush Gupta all maintain a GPA of 3.8-4.0.
Ms. Stevenson was a recipient of the ASU Merit Scholarship and US Army Undergraduate Summer Research Fellowship.
Mr. Simpson graduated from ASU with Honor's and is currently a graduate student at Ohio State University.
Ms. Bing Jiang graduated in May 2013 with a Ph.D. in Biochemistry from ASU

e. Technology transfer—none

f. Accomplishments

This grant enabled us to make considerable progress toward our long-term goal of understanding the physical constraints that limit protein enzyme development.

[1] mRNA Display selection for large globular proteins capable of binding ATP.

We have completed seven rounds of in vitro selection and amplification using mRNA display to isolate synthetic proteins that can fold into structures that recognize ATP with high affinity and specificity. Our starting library (designed and constructed in year 1) contained a common N-terminal de novo evolved ATP binding protein followed by an unbiased random region of 80-amino acids, which was designed to expand the original protein fold by one or more new protein domains. To favor the isolation of stably folded proteins, we included guanidine hydrochloride (GuHCl) in the selection buffer to remove weak or poorly folded proteins from the pool. The concentration of GuHCl was gradually increased over the course of the selection from 0 to 2.0 M. We monitored the selection progress using a cell-based fluorescence assay (see below), which indicated that the relative abundance of stably folded proteins increased significantly from round 0 to round 6 and plateaued in round 7. To determine the diversity of sequences that remained in the pool after seven rounds of in vitro selection and amplification, we cloned 100 representative sequences. As expected, all of the sequences share a common N-terminal sequence that defines the boundary of the parent protein. The C-terminal region of each protein was unique, which may indicate that there are many distinct solutions to the problem of how larger protein structures can emerge from smaller protein domains.

[2] Cell-based fluorescence screen to monitor selection progress and identify soluble protein variants.

We have developed a fluorescence-activated cell-sorting (FACS) assay to monitor the selection progress and identify well-folded and soluble protein variants present in the output of our mRNA display library. The screen is based on a green fluorescent protein (GFP) reporter assay developed by Terwilliger and Waldo (*Nature Biotech*. 1999, 17, 691) that relies on the formation of a productive folding conformation of the upstream analyte when expressed as a fusion protein in *Escherichia coli*. In year 1, we designed and constructed the GFP report vector. This was necessary as we were unable to procure this vector from the Waldo lab, despite several attempts to do so.

Using the GFP reporter vector, we screened populations of sequences from each round of selection for protein folding in *Escherichia coli*. FACS analysis revealed that the starting library exhibited very low fluorescence, similar to the negative (empty vector) control. This is expected as randomly chosen sequences should not fold into stable structures. Analysis of sequences taken from each round of selection showed a gradual increase in florescence, indicating that the selection successfully enriched for proteins with improved protein folding stability.

To evaluate the selection output, we cloned five representative sequences from round 7 into the GFP reporter and assayed each sequence for protein folding in *Escherichia coli*. Two of the five clones, gave a single population of highly fluorescent cells equivalent to our positive control, which contained only parent protein. The other three sequences gave mixed populations of low and high fluorescent cells, indicating that these sequences were adopting partially folded structures.

Encouraged by our FACS analysis, we cloned 64 sequences into the GFP vector and performed a 96-well fluorescence-based assay to identify a subset of sequences from round 7 that were properly folded and soluble. This assay revealed that 18 out of 64 sequences have high relative fluorescence that is reproducible (triplicate). We sequenced the 18 clones and identified 12 sequences with intact open reading frames. These 12 sequences were examined in a standard expression vector (below).

[3] E. coli expression analysis of randomly selected protein clones from round 7.

In addition to the 12 sequences identified in our 96-well protein-folding assay, we chose an additional 23 sequences from round 7 for protein expression analysis. All 35 sequences were cloned into a protein expression vector containing an N-terminal maltose binding protein (MBP) and expressed as C-terminal protein fusions of MBP. The 35 ATP binding protein fusions were expressed in *E. coli* and protein expression was monitored after purification on amylose affinity columns. This analysis led us to identify 4 promising clones that remained soluble as MBP-protein fusions at room temperature for several days. Three of the four proteins yield visible amounts of soluble free protein after cleavage of the fusion protein with thrombin.

[4] Expression optimization by mutagenesis.

To improve protein expression and recovery of the free ATP binding proteins, we made a series of C-terminal truncations to remove amino acid residues that were deleterious. This is a common approach to improving protein solubility and expression. In all three cases, we observed improved expression and solubility when 12 amino acid residues were deleted from the C-terminus.

[5] E. coli expression and purification of three candidate proteins with high solubility.

We are currently optimizing the expression and purification of the three protein candidates so that we can obtain sufficient amounts of protein for biophysical characterization. We are particularly interested in performing an HSQC NMR analysis on each protein to determine the extent to which each protein adopts a discrete protein fold. We are hopeful that at least one of our three proteins is sufficiently well folded to move forward with structure determination studies by NMR and X-ray crystallography.

[6] Protein Characterization by NMR and X-ray crystallography.

The top performing candidate protein identified in our protein solubility screen was selected for structure determination studies by NMR and X-ray crystallography. After considerable effort, conditions were identified that allowed us to obtain milligram quantities of highly pure protein in sufficient concentrations (20 mg/mL) for protein crystallization. However, despite extensive efforts, we were unable to identify suitable crystallization conditions. Because of this result, we shifted our efforts to structure determination by solution NMR. Despite high expression in standard LB media, the protein was extremely difficult to express in minimal media with N-15 labeled ammonia. After several months of screening, we finally identified conditions that allowed us to obtain labeled protein in purified form. HSQC NMR experiments revealed that the evolutionary optimized protein likely adopts a molten globular structure. This unfortunate outcome led us to redesign our library and selection strategy and new efforts are underway to obtain a novel de novo evolved protein with an expanded domain structure.